Chapter 188

December 1987

Massachusetts
Educational
Assessment
Program

The 1986 Massachusetts Educational Assessment Program:

National Comparisons

MASSACHUSETTS DEPARTMENT OF EDUCATION

Publication #15120-22-3000-11-25-87. Approved by Daniel D. Carter, State Purchasing Agent.

ACKNOWLEDGMENTS

The Massachusetts Educational Assessment Program would not have been possible in its present form without the generous support of many individuals and school districts throughout the state. Since its inception in 1985, five Curriculum Advisory Committees have worked to develop the assessment framework, to review the test items, to interpret the results, and to give guidance and support to the entire effort.

We wish to pay special tribute to these committee members, who contributed so much to this process. We are also grateful to the school districts that supported their efforts by granting released time during the school year. The efforts of the one, the support of the other, has resulted in an assessment program that not only reflects the values and academic aspirations of Massachusetts educators, but the most profound thinking in each of the content areas.

MEAP Science Committee

Wayne D. Allen	Leicester High School, Leicester
Mary E. Corcoran	Winthrop Public Schools, Winthrop
Kathleen M. Donnellan	Springfield Public Schools, Springfield
Thomas Heyman	Weymouth Public Schools, Weymouth
Henry Kaiser	Pittsfield Public Schools, Pittsfield
Nicholas Micozzi, Jr.	Plymouth-Carver Regional Intermediate School, Plymouth
Diane Paradis	Assabet Valley Regional, Marlborough
Edward Pratt	Longmeadow High School, Longmeadow
F. Paul Quatromoni	Dedham High School, Dedham
Mary Ann Sudolnik	Turkey Hill Middle School, Lunenburg
John A. Tyrell	Boston Public Schools, Boston
Michael Zapantis	Beverly High School, Beverly

MEAP Social Studies Committee

Virginia H. Ahart	Hampshire Regional High, Easthampton
Charles B. Cormier	Robinson Elementary School, Mansfield
June R. Coutu	King Phillip Regional High, Wrentham
Hazel C. Donnelly	Durfee High School, Fall River
Patricia Dye	Plymouth-Carver Regional Intermediate School, Plymouth
William M. Dyson	Glenbrook Middle School, Longmeadow
Bernadette Golden	Lincoln School, Winchester
John M. Hassan	Daley Junior High School, Lowell
Henry Hicks	Pollard Middle School, Needham
Edward Lerner	Newton Public Schools, Newtonville
Hugh T. McDonagh	Wheatley Middle School, Boston
Debra J. Miller	Belmont Street Community School, Worcester
Garry Murphy	Greater Lowell Regional Vocational Technical, Tyngsborough
George S. Perry, Jr.	Massachusetts Department of Education
Jane W. Rowe	Provincetown Elementary School, Provincetown
Don Salvucci	Brockton High School, Brockton
Ronald Toleos	Ipswich High School, Ipswich
George G. Watson, Jr.	Winchester High School, Winchester
Mary Ann Wolff	North Reading High School, North Reading

— Table of Contents —

CHAPTER 1

Highlights

This report compares information from the 1986 Massachusetts Educational Assessment Program to similar data from the National Assessment of Educational Progress. Many of the important findings of the Massachusetts Educational Assessment Program already have been published. This report supplements those earlier reports, adding in the important supplemental information of how Massachusetts schools and students measure up to national performance.

Three content areas were measured by both assessments: reading, mathematics, and science. The National Assessment of Educational Progress has not yet released its reading data, and therefore comparison of Massachusetts to the nation could only be made in mathematics and science. In both content areas, at all three grades, the achievement levels of Massachusetts students were above those of students nationwide.

Mathematics was an area of particular strength. More importantly, Massachusetts students did best on questions that challenged their ability to extract information and manipulate it to solve problems. This is exactly the skill that most curriculum specialists would say is the one that should be emphasized. While Massachusetts students did well on routine problems, it was their exemplary performance on more challenging questions that suggests that the appropriate instructional focus is being applied in Massachusetts.

Students also did well in science, although not quite as well as they did in mathematics. At all grade levels they performed better than the national sample in the area of life science, particularly on those questions that dealt with health or practical applications. However, they showed a relative weakness in their understanding of physical sciences, particularly at the eleventh grade level. They showed particular strength in their knowledge of and ability to use methods of scientific inquiry. This result is somewhat surprising in light of questionnaire results that show that Massachusetts students are somewhat less likely than their counterparts nationwide to have an opportunity to do laboratory work, either in small groups or individually.

Background factor questions on the questionnaire indicated that Massachusetts students are not as different as one might think in terms of educational advantage in the homes from which they come. Their parents have levels of education that are not much different from national averages, and the proportion of them who speak a second language at home is virtually identical to nationwide averages. Nonetheless, the educational aspirations of Massachusetts students far exceed those of students nationwide. Massachusetts students are far more likely to plan to continue their education, and are far more likely to be taking college preparatory courses in high school.

One final important finding concerns the aging of the teaching population. While teachers nationwide are growing older, because fewer young people are joining the teaching force, this trend is far more pronounced in Massachusetts. This phenomenon certainly has far ranging consequences for the long term, and is an issue that must be addressed soon.

CHAPTER 2

Introduction

The Massachuetts Educational Assessment Program was established in 1985 under the provisions of the School Improvement Act of 1985 (Chapter 188). The purpose of the assessment program is twofold: to furnish information to improve curriculum and instruction, and to compare schools and districts at the state and national levels. In the spring of 1986 virtually all students in grades 3, 7, and 11 were tested in reading, mathematics and science. The results of that testing were reported to schools and districts in the autumn of 1986. Those test reports have been followed by a series of curriculum reports, which have considered the assessment results in the light of curriculum improvement. The purpose of this report is to examine results in contrast to national achievement, as mandated by law.

In order to produce national comparisons, the Massachusetts tests made extensive use of materials from the National Assessment of Educational Progress (NAEP). NAEP was chosen as the vehicle for the national comparisons because its purposes are consistent with those of the Massachusetts assessment (i.e., it is specifically designed to test curriculum rather than to distinguish among individual students), and it provides accurate national data. To make comparisons to NAEP national results as valid as possible, the Massachusetts assessment selected the same grade levels, administration times, and subjects to be tested as NAEP.

Two appendices are included with this report. The first provides a short overview of the Massachusetts Educational Assessment Program. The second describes the differences between the Massachusetts assessment and NAEP, and provides some detail on the attempts that were made to make the data from the two assessments as comparable as possible.

It is somewhat ironic that the legislation that created the mandate for the Massachusetts assessment to provide national comparisons also contained another provision that made it impossible to meet that mandate with a high degreee of precision. As noted above, NAEP was chosen as the vehicle for making national comparisons in part because it is the only source of truly accurate national data. However, NAEP tests are administered under a set of rules and conditions that are different from the Massachusetts assessment. NAEP data are reported at the regional and national levels only; no local results, even at the state level, are made public. On the other hand, Chapter 188 mandated that the Massachusetts assessment results be made public at the school and district level. The difference in reporting requirements had many effects, some direct and some indirect.

The direct effects are rather obvious. The motivating factors for both examinees and test administrators are quite different when results will be reported locally than when they will not. While no specific research has been done that would shed light on the impact of this effect, it is reasonable to assume that an increase of performance caused by local reporting could be as much as two or three percentage points. An example of the impact of motivation can be found in the California Assessment Program. Before the second year of that testing program, the department decided to add a space to the answer sheet so that twelfth grade students could record their names. Statewide performance rose two to four percentile points that year—the first year of increased performance at that grade in almost a decade.

While the indirect effects of local reporting are less obvious, they are far more numerous. For example, while it might have been possible to use the NAEP tests intact if no local reporting had been done, it was necessary to revise those tests to make them appropriate for Massachusetts once the decision to have local reporting was made. Another example of the indirect effects that local reporting had on test design was the timing of the tests. NAEP administered its tests in 16-minute, timed blocks which contained too much material for many students to complete. While this practice sufficed for NAEP's purposes, the testing in Massachusetts would have been seriously criticized had the same administrative methods been used. For two of the three blocks used in the Massachusetts tests, students were given all the time they needed to complete the test.

Another difference that the Chapter 188 mandates led to was the population tested. Chapter 188 called for the testing of all students, with limited exceptions. NAEP also called for the testing of all students, except for those "deemed unassessable by the school authorities." This meant, in some cases, that students who would have been tested for NAEP were excluded from the Massachusetts assessment. Also, in part because there would be local reporting, the Massachusetts assessment provided for three weeks of makeup testing in order to insure that all students who could be tested would be tested and included in their school's report. Over 97 percent of students who were not special needs or limited English speaking were tested in the Massachusetts assessment. This is a far higher percentage than is true for NAEP, where students who are absent from school during the period of time that the NAEP administrator is in their school do not get tested.

As a result of all these effects, the comparisons provided in this report must be interpreted with caution. A more detailed description of the issues involved is provided in Appendix B.

Finally, this report does not include comparisons of reading performance. Although NAEP passages and questions were included in the Massachusetts assessment at all grade levels, the National Assessment finds that it is unable to produce data at this time. Consequently, the achievement comparisons are confined to the areas of mathematics and science.

CHAPTER 3

Achievement

As noted in the previous chapter and Appendix B, there were several factors that led to differences between the administration of the tests in Massachusetts and NAEP. While there were hundreds of test questions at each grade level that were used by both assessments, only a limited number of questions, and only a subset of the data available on each question, could be considered reasonably comparable. Test items are reported in this section only when they were responded to by at least 95 percent of the national sample. Also, only data provided by students who took the test items in the same block (first, second, or third) in both assessments are used. Thus, while there are many cautions to be considered in the interpretation of these data, as noted in Appendix B, the information reported in this chapter is the carefully selected subset that can be most fairly deemed comparable.

The first item of interest is the overall results. Those are provided in Table 1.

Table 1

Overall Differences between NAEP and Massachusetts Assessment Test Results

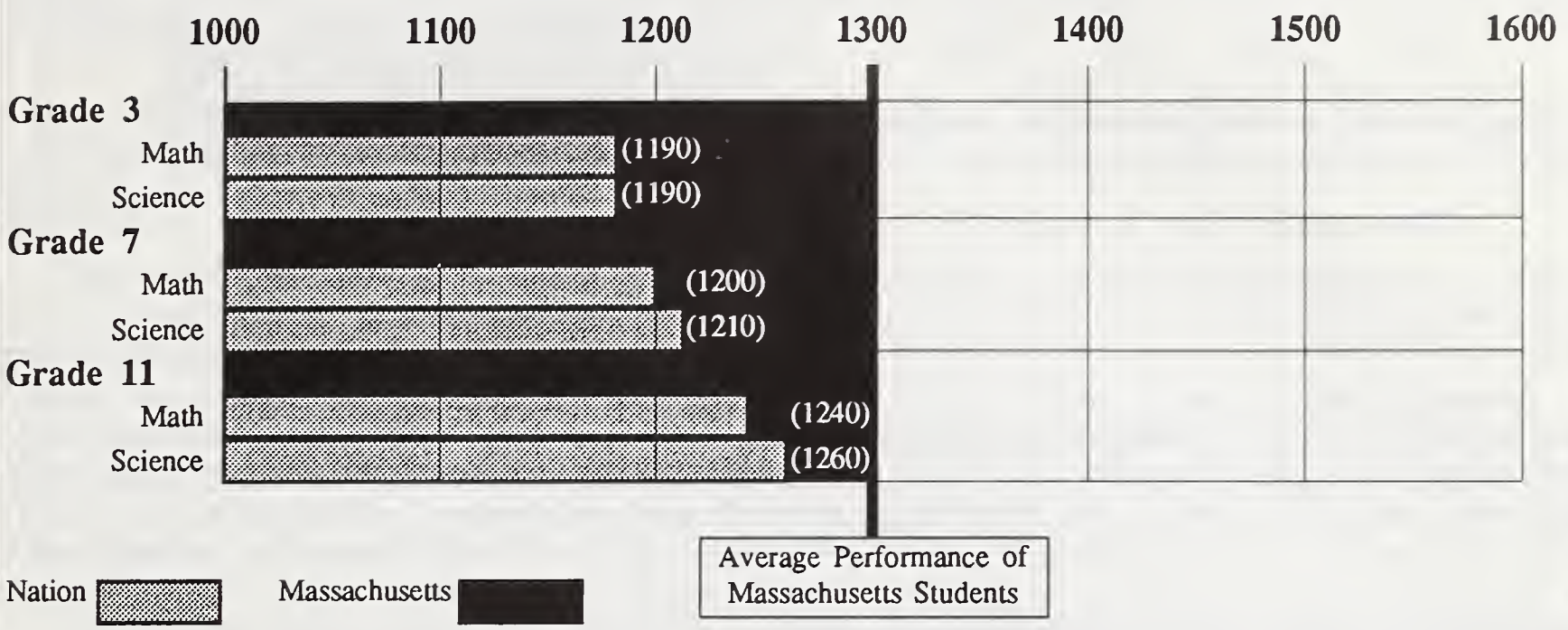
Grade	Content Area	Number of Questions with Comparable Data	Average percentage of correct responses in Massachusetts over NAEP
3	Mathematics Science	29 25	7.5 7.9
7	Mathematics Science	36 63	8.3 5.8
11	Mathematics Science	74 90	4.3 2.6

As is shown in Table 1, students in Massachusetts scored higher than their counterparts nationwide at every grade in both content areas. The advantage shown by students at grade 11 was smaller than it was in the other two grades. At the upper two grades, the scores of Massachusetts students, relative to national performance, were somewhat higher in mathematics than in science.

When the reporting for the Massachusetts assessment was done, results were given as scaled scores developed specifically for that program. This scale was established with a range of 1000-1600, a mean of 1300, and a standard deviation of 100. Table 2 provides the estimated performance of students nationwide on the Massachusetts scale.

Table 2

Estimated Performance of Students Nationwide on the Massachusetts Assessment Score Scale



Using this table, then, a school that had scored higher than 1190 on the Massachusetts Educational Assessment Program in grade 3 mathematics could assume that its students would score higher than a national sample. Since over 80 percent of the schools statewide had scores on the Massachusetts assessment that were 1200 or higher at each grade, this means that the vast majority of schools throughout Massachusetts, especially at grades 3 and 7, have average test scores above national averages on the common set of items tested.

Perhaps of even more interest than overall results, however, is the performance of students on subsets of questions. Average performance was very good; however, that average result comes about from some areas that were not as strong, while others were outstanding. The rest of this chapter will be devoted to looking at those results, grade by grade.

It should be remembered when reading these results that the discussion is in terms of relative, rather than absolute, performance. In our state and focused reports, we have used a different standard. In those reports, we compared student performance on items to what our curriculum area committees believed *should* be achieved. In the present instance, we are comparing performance to what other students *do* achieve.

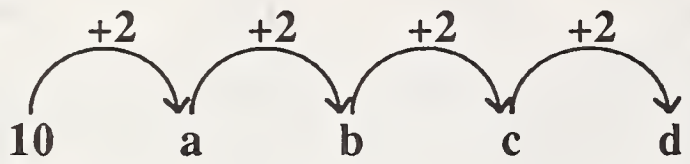
Mathematics

Grade 3

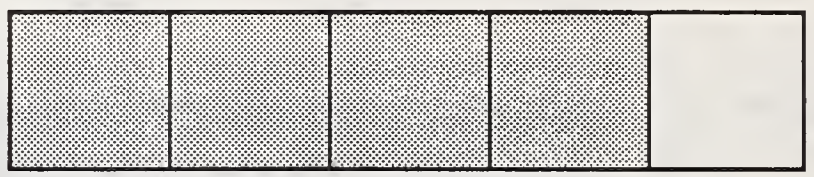
Of the 29 grade 3 mathematics test questions on which national comparable data are valid, Massachusetts students score higher than NAEP results on 25 of them. The range of performance was -6 (that is, on one item, students nationwide scored 6 percentage points higher) to +16 (that is, students in Massachusetts scored 16 percentage points higher than the national sample).

Although the students performed generally well on most of the comparison items, the items on which their performance clearly outstripped the national sample indicated an advantage in the understanding of operations and the use of mathematical reasoning skills.

The largest difference between the two groups was in response to a word problem in which students were required to recognize the correct operation when information was presented in a verbal form. However, substantial differences in achievement were also found when students were presented with graphical representations of operations as, for example, when they were asked to complete an addition operation that was portrayed on a number line or to identify a number sentence that was represented by a chart.



What number does 'd' represent?



Which answer tells about the shaded boxes above?

Answer: ☐ + 1 = 5

The unifying theme among this set of items was the need to recognize the underlying operation that was portrayed, either visually or verbally. Students' success in this recognition and their ability to move freely between different representations of operations implies a more intuitive understanding of these operations than is true of the national sample. That these results reflect instruction is suggested by the third grade teacher questionnaire in which teachers were asked to respond to sample test items. Although none of these NAEP items were used, almost all teachers indicated that their students had been taught to recognize the underlying operational sentence in verbal problems. One may assume that this must be true for visual contexts, as well.

However, students also performed better than the national sample when the demands of the task were greater. For example, in some items students were not only asked to recognize and extract information presented in graphical form, but to analyze that information in order to solve problems, as in the following question.

> Given a graph representing the number of oranges, lemons and grapefruit picked on each day of the week, students were asked, "On which day were more boxes of lemons picked than either boxes of oranges or boxes of grapefruit?"

Although Massachusetts students were not necessarily better than the national sample in reading the graph, they outscored the national average by ten percentage points when asked to apply logical relationships to the information contained in the graph.

Massachusetts students also scored especially well on two questions dealing with money. On a question that asked them to select three nickels and one dime as being the same amount of money as a quarter, Massachusetts students scored above the national average by 15 percentage points. Another question combined elements of word problems, estimation, and money; it asked for the smallest bill that would be enough to buy four baseballs at $4.70 each. This, of course, was a difficult question for third graders, but students in Massachusetts scored eight percentage points higher than the national average.

Grade 7

Of all the grade levels and content areas, Massachusetts students did best, relative to national performance, in grade 7 mathematics. Of the 36 questions on which comparable data were available, Massachusetts students scored higher than their counterparts nationwide on 33 of them, by as much as 17 percentage points. On the three questions on which they scored below national results, they did so by only one percentage point.

As was true for the third grade, students in this grade did relatively well on questions involving logic and word problems, such as one which described relative ordering. For example,

G is first
B is next
D is between S and B
Who is last?

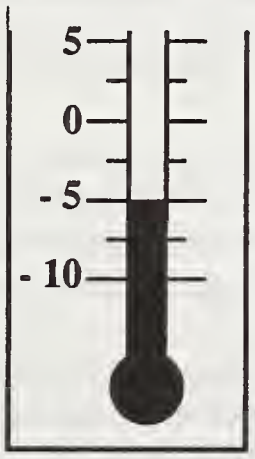
Sixty percent of the students in Massachusetts correctly deduced that S must be last, compared to 44 percent of students nationwide. On a similar problem requiring deductive logic, Massachusetts students scored above the national results by 15 percentage points.

Also as was true for the third graders, the seventh grade students did especially well on questions involving money. On the two word problems that required knowledge of money, Massachusetts students outperformed the national averages by 11 and 12 percentage points.

Students did well, although not at quite the level of the above areas, in questions involving graphs. In the six questions associated with two graphs, students in Massachusetts scored between six and eight percentage points better than the national averages.

Compared with their superior performance in other areas, students in Massachusetts did not do nearly so well with questions testing knowledge of units and metric measurement. Only 43 percent knew that one kilogram equals 1,000 grams (compared to 44 percent nationwide), and 69 percent selected "7 feet" as the length closest to the height of the classroom door (compared to 70 percent nationwide).

The following question, which was administered to both third and seventh graders, had an interesting result.



	Grade 3	Grade 7
Massachusetts	74	90
Nation	58	87

Students were asked to read a thermometer that showed five degrees below zero. At grade 3, 74 percent of the students in Massachusetts answered the question correctly, compared to just 58 percent nationwide. This gap of 16 percentage points represented the greatest distance between students in Massachusetts and the national sample. At grade 7, 90 percent of the students in Massachusetts answered the question correctly (an increase of 16 percentage points over the third graders' performance), but now that was just three percentage points greater than the national average. Performance nationwide went from 58 percent to 87 percent between those two grades. It may well be that students in Massachusetts simply learn about negative numbers sooner, but by the seventh grade have little or no edge in this area; however, it is just as likely in this case that by the seventh grade, this question is just so easy for everyone that it is virtually impossible for students in Massachusetts to show their edge if it exists.

Grade 11

Of the 74 mathematics questions on which comparable data for eleventh graders were available, Massachusetts students scored above their national counterparts on 63 and equalled them on four more. Of the seven questions on which the national sample scored higher than Massachusetts students, on only one question was the gap more than two percentage points. However, the close match of the performance of the two groups was equally evident

at the other end of the distribution: on only two questions did the performance of Massachusetts students exceed that of the national group by more than 10 percentage points.

Given this relatively level performance, it is difficult to discern particular areas of strength or weakness. Certainly, however, the preponderance of questions on which the performance of Massachusetts students was relatively strong were complex word problems. A typical example would be the verbal equivalent of the following type of problem:

$$A = \frac{X}{4}$$
$$A = 5$$
$$X = ?$$

Seventy percent of the students in Massachusetts were able to recognize the underlying structure or the verbal equivalent and obtain the correct solution, compared to 60 percent in the NAEP sample.

Another example of the superiority of Massachusetts students in solving non-routine problems was a problem that showed students a set of boxes, each of different weights. Students were to select those whose total weight equalled half the weight of the entire set. At both seventh and eleventh grade, Massachusetts students far exceeded the national averages; by 17 percentage points at grade 7 and seven percentage points at grade 11.

Performance was also particularly strong on a non-routine problem involving pie charts that was administered to students at both grades. Students were provided with a pie chart that showed a recommended plan for distributing a food dollar. They were then asked to select the correct policy recommendation that was portrayed by the chart.

At grade 7, students in Massachusetts scored above their national counterparts on this question by 15 percentage points; at grade 11, by six points.

Students also did well on questions involving money, just as they had done at the earlier grades. On the same item noted as a relative strength for the third graders (determining the smallest bill necessary to buy four baseballs at $4.70 each), the eleventh grade students also scored eight percentage points above the national average.

While there were no glaring weaknesses in the performance of the eleventh graders, they showed the same lack of ability to use fractions and percentages that seems to be true of students and adults nationwide. The worst performance of Massachusetts students, relative to national results, was on a question that required changing a batting average expressed as a decimal to a percent. Only 59 percent of the Massachusetts students could do that correctly, as opposed to 64 percent of the national sample. Massachusetts students also scored below the national average on a problem that asked them to recognize that the correct algorithm for the multiplication of fractions. More students selected the incorrect answer of than chose the correct answer. Massachusetts students also scored comparable to the national results on two more similar items involving percentages: "To find 125% of a number, you can multiply by 1.25"; and "1.00% is not the same as 100%." On both items, fewer than three students in five were able to answer the item correctly, both in Massachusetts and in the national sample.

Science

Grade 3

In grade 3 science, 25 questions remained on which national comparisons could be made. On all 25 questions, Massachusetts students scored above the national averages, ranging from a low of 1 percentage point to a high of 15.

As will be shown in more detail in the discussion of science at the upper grades, students in Massachusetts are able to apply the principal of conservation of matter much better than students nationwide. Two items particularly illustrate this at grade 3. In one item, a brick was placed on a scale in three different positions. Students were asked in which position the brick would weigh the most. The correct answer, "The brick will weigh the same in all three positions," was selected correctly by 13 percent more students in Massachusetts than in the national sample. Similarly, Massachusetts third graders scored above the national results by seven percentage points on a question that showed a boy holding a balloon and then squeezing it. The correct answer here, of course, was that the weight of the balloon stays the same no matter what shape the balloon is forced into.

Also as in the upper grades, however, Massachusetts students tended to carry the concept of conservation of matter beyond its limits. Shown three glass tubes of varying diameters, but all filled to the same level, only one percent more students in Massachusetts were able to correctly identify the fact that the tube with the widest diameter would have the most water. More students in Massachusetts than nationwide incorrectly thought all tubes would have the same amount of water.

Third graders did particularly well in life sciences. With one exception, Massachusetts third graders surpassed national results by at least seven percentage points on life sciences questions. Of the eight questions on which Massachusetts students scored at least ten percentage points higher than national averages, five were life sciences questions. As will be seen shortly, this too was a result that carried over to the upper grades.

Grade 7

There were 63 science questions asked of grade 7 Massachusetts students for which data from their peers nationwide were available. While the Massachusetts students scored higher on 53 of them, this will be the first area discussed in which the performance of the national sample was significantly higher than Massachusetts for a substantial number of items.

With one surprising and unexplainable exception, students in Massachusetts did relatively well in questions related to conservation of matter. Two questions especially showed this. One question showed two objects, each made of four identical cubes. One object was made by aligning the cubes in a row; the other was made by connecting the cubes in two rows of two each. In Massachusetts, 63 percent of the students realized both objects would have the same weight and volume, while just 47 percent of the students nationwide answered the question correctly. Another question showed a sealed can filled with crushed ice. Students were expected to know that after the ice melted, the can would continue to weigh the same. Again, students in Massachusetts scored above the national group by 15 percentage points. The surprise came on a question that third grade students had done very well on. That question, discussed above, was the one in which the brick was placed on a scale in three different positions. At grade 3, 60 percent of the Massachusetts students answered that question correctly, a result 13 percentage points better than the national average. That is a result consistent with the finding above. However, when that same item was taken by seventh graders in Massachusetts, 67 percent answered it correctly, which was approximately the same as the national result.

Students also did very well in the aspects of life science dealing with human health in practical ways. For example, they scored 13 percentage points above the national results on identifying the effect of alcohol on the body, and nine percentage points above on two similar questions (what a doctor is doing when the doctor rubs an unmedicated cotton swab on the back of your throat and both animals and plants perform respiration). However, on life sciences questions that would most likely be learned in an academic environment, students did not do nearly so well. Massachusetts students scored six percentage points below the national group in knowing that red blood cells carry most of the oxygen to the body, and five percentage points above the nation on being able to order a group of organisms from simplest to most complex.

Students in Massachusetts did extremely well on questions related to the methods of scientific experimentation. For example, two questions asked how the factors causing plant growth could be isolated. On one question, the Massachusetts students scored 13 percentage points above the national average, and 12 points on the other. On a

question that asked, "Cooked vegetables are often not a nutritious as the same kinds of vegetables uncooked. How could you test this statement?", students in Massachusetts also scored higher than the national sample by 13 percentage points. Several other examples of high performance by Massachusetts students in this area make it clear that it is a solid relative strength of instruction in the state.

On the other hand, performance on questions dealing with physical science were not nearly so good. Only 44 percent of the students in Massachusetts knew that ringing a bell would not be a satisfactory way to communicate on the surface of the moon (as opposed to other methods that used transmission of light waves), compared to 52 percent of NAEP's sample. They also scored eight percentage points below the national sample on a question that involved the relative effect of force on mass, with only 50% of the students giving the correct response.

Grade 11

While eleventh graders in Massachusetts scored above national averages in this area, grade 11 science was the grade level and content area combination that showed the poorest performance of Massachusetts students relative to national performance. Of the 90 questions asked on which comparable data were available, the national sample scored higher than Massachusetts students on 23 of them, and by as much as 11 percentage points. Students at grade 11 showed many of the same strengths and weaknesses as did the seventh grade students.

Just as in the seventh grade, the eleventh grade students did very well in questions related to the conservation of matter. For example, one question. For example, one question presented the following situation:

> In a laboratory, A and B are mixed together, producing two other chemicals, C and D. Given the quantities of A, B and C, how much D is produced?

Forty percent of the students in Massachusetts correctly answered this question, compared to 28 percent in the national sample. Also, the question about the ice melting in the can that was discussed in the section on grade 7 science also was asked of eleventh graders. Students at this grade also scored above the national sample, this time by six percentage points.

If there was a problem in their knowledge of conservation of matter, it was that they tended to carry the principle beyond its logical extension. For example, students were presented with two situations. In one, grain was grown as food for humans. In the other, it was used as feed for livestock which, in turn, were slaughtered for their meat. Only 27 percent of the eleventh graders in Massachusetts knew that the grain would provide sufficient food for more people than the meat, a result that was two percentage points below the national sample. Almost half of the Massachusetts students thought that the meat and grain would each provide sufficient food for the same number of people, which suggests that many of them may be using the principle of conservation of matter in inappropriate situations.

Again as in the seventh grade, eleventh graders in Massachusetts did well on questions related to methods of scientific inquiry. On two related questions dealing with a controlled experimental condition, Massachusetts students scored above the national average on both. One question required them to determine the variable under study, and the second asked them to select the most valid conclusion based upon the evidence. Similarly high performance was observed on a question that involved determining the logical flaw in an informal experiment (two variables were uncontrolled, rather than one). In yet another question in which students were asked to recognize that the result of a natural experiment indicated only the possibility, rather than the certainty, of a causal relationship between two variables, Massachusetts students scored above the national average by five percentage points.

Eleventh graders also performed similarly to the seventh graders in that they did very well in questions related to human health and practical aspects of life science. They scored higher than the national sample by more than ten percentage points on each of four questions related to this area. One question involved understanding the cause of "jet lag." Another required them to identify sodium bicarbonate as a substance useful in relieving the discomfort of an acid stomach. The third question required them to know that when a few individuals of a species survive an attempt at extermination, that illustrates one of Darwin's key ideas: that there is a variation among individuals within a species. The final question combined both elements of knowledge of the limitation of experimental evidence

and human health. Students were asked what scientific validity supported the Surgeon General's warning that "cigarette smoking is dangerous to your health." Students, both in Massachusetts and the nation, tended to select only two possible answers: one that was correct, citing statistical data show a higher incidence of respiratory and cardiac disease in smokers than in nonsmokers; and one that implied more causative evidence linking smoking and cancer than actually exists at present. The correct choice was selected by 47 percent of the students in Massachusetts, compared to 35 percent of the national sample.

Students in Massachusetts, however, tended to score poorly in questions related to physical science, just as they had done in the seventh grade. Of the 23 questions on which Massachusetts students scored below the national sample, 11 measured knowledge of physical science. In an earlier section, it was noted that seventh graders did not do well, compared to national results, on the question about not being able to communicate on the moon by ringing a bell. The eleventh graders did even worse, relative to their peers, than had the seventh graders, scoring ten percentage points below the national average. They also scored six percentage points below the NAEP sample on a question that involved knowing that a plane must release its cargo before reaching its target. Another physical science question on which the Massachusetts eleventh graders scored below the national average was one that required them to know that a copper wire would increase in length when heated. Thus, the edge in practical knowledge that seventh graders showed on their test results did not carry through to the eleventh graders.

CHAPTER 4

Questionnaires

In addition to the extensive sets of achievement questions that were asked of students in both the Massachusetts and the national assessments, several other questions were asked. These questions dealt with the backgrounds of the students and their attitudes. Beyond those student questionnaires, an additional set of questionnaires was developed for the principals to complete in both assessments. Many questions in both the student and the principal questionnaires were the same in both the Massachusetts and national assessments, providing another avenue through which similarities and differences between Massachusetts and the rest of the nation can be explored. In this chapter, the student questionnaire results will be discussed first, followed by the principal questionnaires.

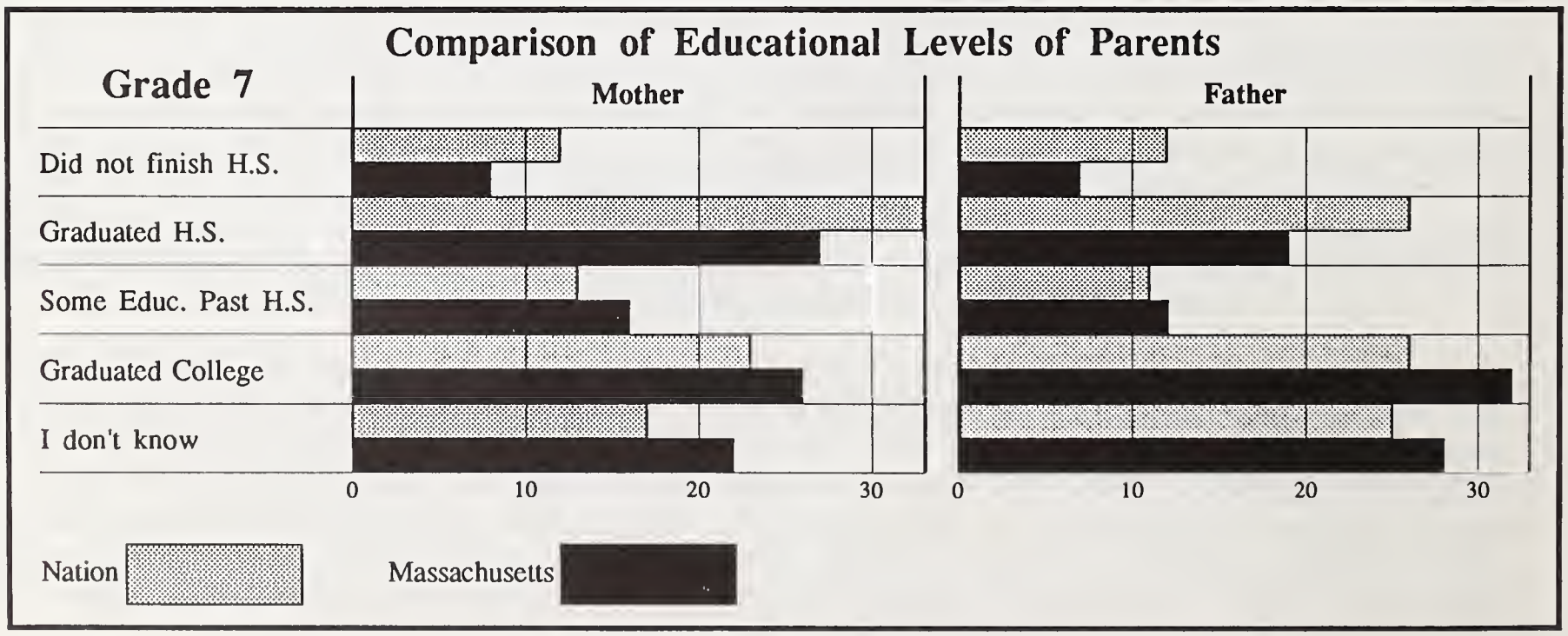
Student Questionnaires

Background Factors

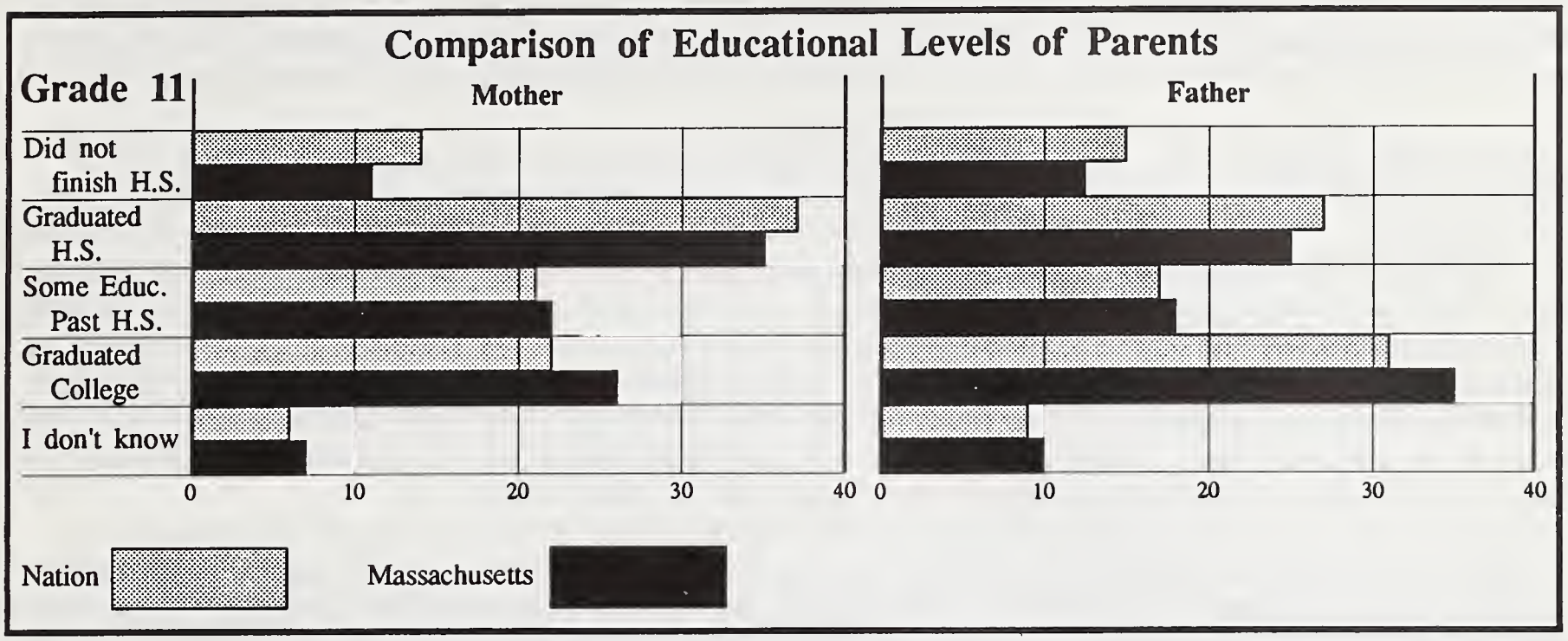
The Massachusetts assessment thought it was important to read the questionnaire items to third graders. As a consequence, no matrix sampling of the student questionnaires was attempted at this grade. As a result, very few questionnaire items could be asked, and only a very limited set were selected from NAEP. The few that were jointly administered, however, yield some interesting results.

The one questionnaire item used at grade 3 in both assessments that comes under the category of "Background Factors" was the one that asked students whether they had gone to nursery school or Headstart before entering kindergarten. Sixty one percent of students in Massachusetts answered affirmatively, compared to 55 percent nationwide.

Several of the background questions that were asked at grades 7 and 11 indicate that students in Massachusetts have home environments that, on the average, are much like what is found across the country, although they are somewhat more advantaged on some items. At both grades, students were asked about the level of their parents' education. Nationwide, 36 percent of the seventh grade students indicated that their mothers had attended school beyond high school, compared to 42 percent of the students in Massachusetts. Similar statistics on father's education were 37 percent for the national sample and 44 percent for Massachusetts. However, Massachusetts students were

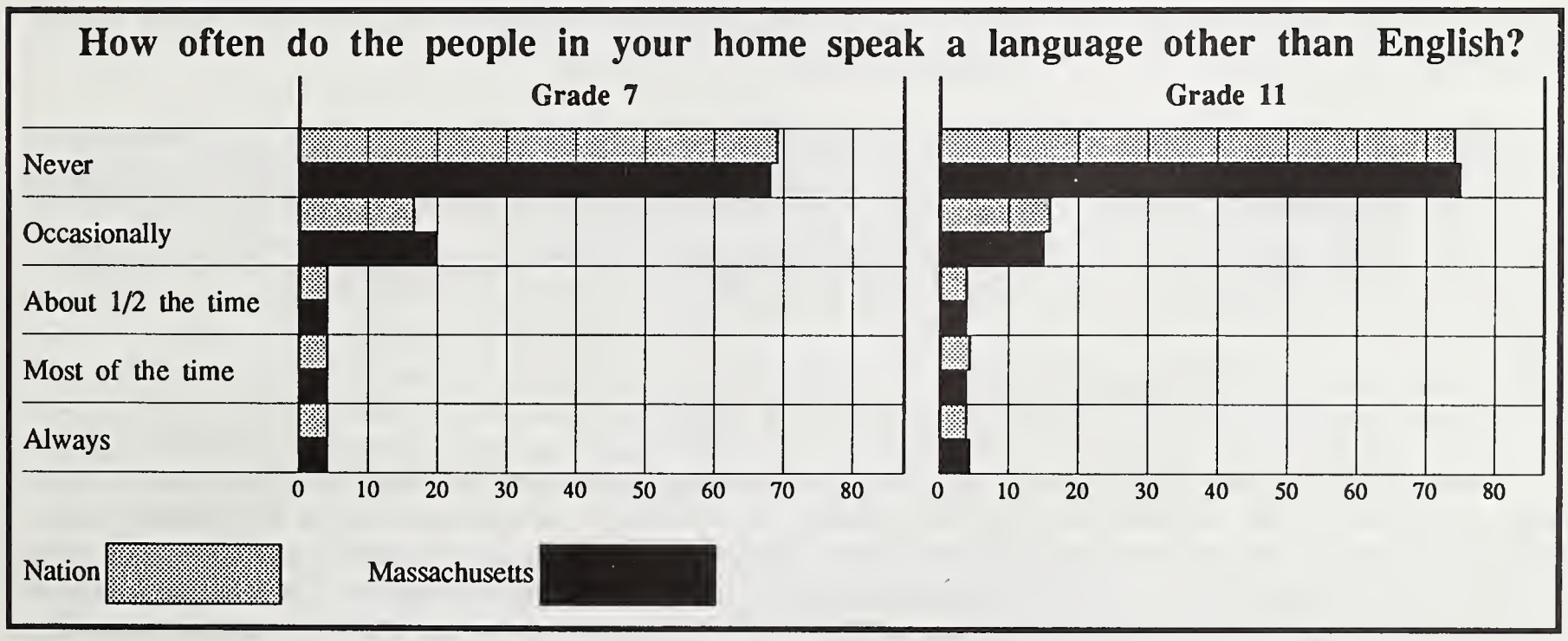


less likely to know their parents' educational level. Twenty two percent of the seventh grade students in Massachusetts did not know their mother's educational level, compared to the national sample of 17 percent. For father's education, the figures were 28 and 25 percent, respectively.



The questions about parent education indicated an even closer match between Massachusetts and the country at the eleventh grade than they did at the seventh grade. As opposed to the seventh grade results, almost equal percentages of students in Massachusetts and the national sample answered "I don't know" to the question about both mother's and father's education, so the data are more clearly comparable. Twenty-six percent of the eleventh graders in Massachusetts indicated that their mothers had graduated from college, compared to 22% of the national sample. The similar statistics for the two groups on father's education were 35% and 31%, respectively. Thus, the parents of students in Massachusetts had somewhat higher education, but there was not a large difference between the two groups.

Also, there was no difference at all at either grade level between the responses of the Massachusetts students and the national sample to the question about speaking a language other than English at home. While Massachusetts is thought of as a state in which second languages have a significant impact, the questionnaire results indicate that the extent of use of languages other than English is no different in Massachusetts from the remainder of the country. Whatever degree of bilingualism exists in Massachusetts is virtually identical to a national cross-section.



Given the fairly equal levels of education between parents in the state and parents nationwide, it is not particularly surprising that the amount of reading done by families in Massachusetts, as perceived by students, is similar to national averages. While a slightly higher percentage of seventh grade students in Massachusetts reported that the people in their home read a newspaper, magazine or book every day, it also was true that an even higher percentage than the national sample reported that such materials were "hardly ever or never" read in their homes. Eleventh graders also reported the same lack of difference in home reading habits of their family that the seventh graders did.

Student Practices

Television watching has been related to test score performance in many studies. Children who watch several hours a day of television tend to have poorer test results than those whose watching is more moderate. Students in Massachusetts report watching less television than their peers nationwide. In Massachusetts, 38 percent of the seventh graders reported watching two hours or less of television each day. In the national sample, only 28 percent of those responding to the question watched that little television. The difference between the two groups was even greater at the eleventh grade. At that grade, 66 percent of the Massachusetts students reported that they watch two hours a day or less of television, as opposed to 44 percent of the NAEP group.

Homework is another variable often looked at when reviewing questionnaire results for keys to differential test performance. Seventh graders were asked about the amount of science homework they do, and students in Massachusetts answered almost exactly the same as their counterparts nationwide. Similar results were found at grade 11. Of the slight difference that did exist, Massachusetts students were more likely to get homework, and spend slightly longer on it.

The above information shows that there was little difference between the national sample and Massachusetts students in several important areas: parents' educational levels, rates of bilingualism, amount of television watching, and amount of homework. However, there were clearly different answers given on one crucial area asked at grade 11—the educational aspirations of students. Sixty-one percent of the students in Massachusetts described their high school program as "academic or college preparatory," as opposed to 48 percent in the national sample. Also, 60 percent of the Massachusetts students described their current English course as "advanced placement or honors" or "college preparatory," as opposed to 43 percent of the national sample. This confirms the findings of a 1985 report of the Department (*Course-Taking Among Massachusetts High School Students*), which found that Massachusetts students average almost one more academic course per year than their national counterparts and far exceed the participation rate of students nationwide in science and mathematics courses taken.

The present survey found that 89% of the eleventh graders in Massachusetts reported having taken a general science course, as opposed to 76 percent of the NAEP sample; the comparable statistics were 85 versus 84 for biology, 54 versus 36 for chemistry, and 15 versus 10 for physics.

The answers to questionnaire items related to mathematics also reinforced the observation that Massachusetts students have higher educational aspirations. A smaller percentage of Massachusetts students have taken general business or consumer mathematics (31 percent in Massachusetts, versus 49 percent in the national sample), but a larger percentage have taken college track mathematics courses: pre-algebra, 65 percent versus 57; first-year algebra, 82 versus 78; second-year algebra, 57 versus 45; geometry, 71 versus 59; and trigonometry, 20 versus 18.

Student Perceptions about School

Grade 3. Third grade students were asked about classroom practices regarding reading. While some caution needs to be taken in the interpretation of the data because of a far larger non-response rate in the national sample (perhaps due to problems that were eliminated by reading the questionnaires to third graders in Massachusetts), there remain some very strong findings. In Massachusetts, 48 percent of the third graders reported that they read out loud in reading class almost every day, while nationally, only 39 percent responded that way. Perhaps even more revealing

is the fact that 21 percent of the national sample said they never read out loud in reading class, compared to only four percent of the students in Massachusetts.

Silent reading practices at grade 3, however, were very similar across the two groups, although silent reading is a slightly more common occurence in Massachusetts. Six percent of the national sample reported that they never read silently on their own in school, compared to two percent in Massachusetts.

Grade 7. One substantial difference between Massachusetts students and the national sample was in answer to the question, "How often do you look forward to your science class?" Eight percent more students in Massachusetts answered, "almost every day," and eight percent fewer said, "never" than the national sample. Certainly, then, Massachusetts science teachers are doing more to keep their students interested in science class than is true for the remainder of the country.

One reason for science class being more interesting to students may be what their teachers are doing in class. Almost twice the percentage of seventh grade students nationwide (40 vs. 21 percent) reported that they read their science textbooks in science class almost every day.

On the other hand, the answers to several other questions painted a somewhat different picture of science instruction in Massachusetts. Two-thirds of the students in Massachusetts, compared to just half in the national sample, reported that they "hardly ever or never" do experiments of laboratory work by themselves. Also, Massachusetts students were more likely to hear a lecture for their science class. Thirty-nine percent of the Massachusetts seventh graders reported that their teacher lectures almost every day, as opposed to 31 percent in the national sample.

In mathematics, the pattern of responses about classroom activities was generally similar in Massachusetts and the remainder of the country. One substantial difference was that 78 percent of the students in Massachusetts reported that they "hardly ever or never" work in small groups in math class, as opposed to 63 percent of the NAEP sample.

Grade 11. Few questions were asked about practices in English classes that were common to the two assessments, but one that was indicated a clearly positive result for Massachusetts teachers. When asked how often their teachers show them how to outline main ideas, 43 percent of the Massachusetts students answered at least several times a week, while only 25 percent of the national sample did so.

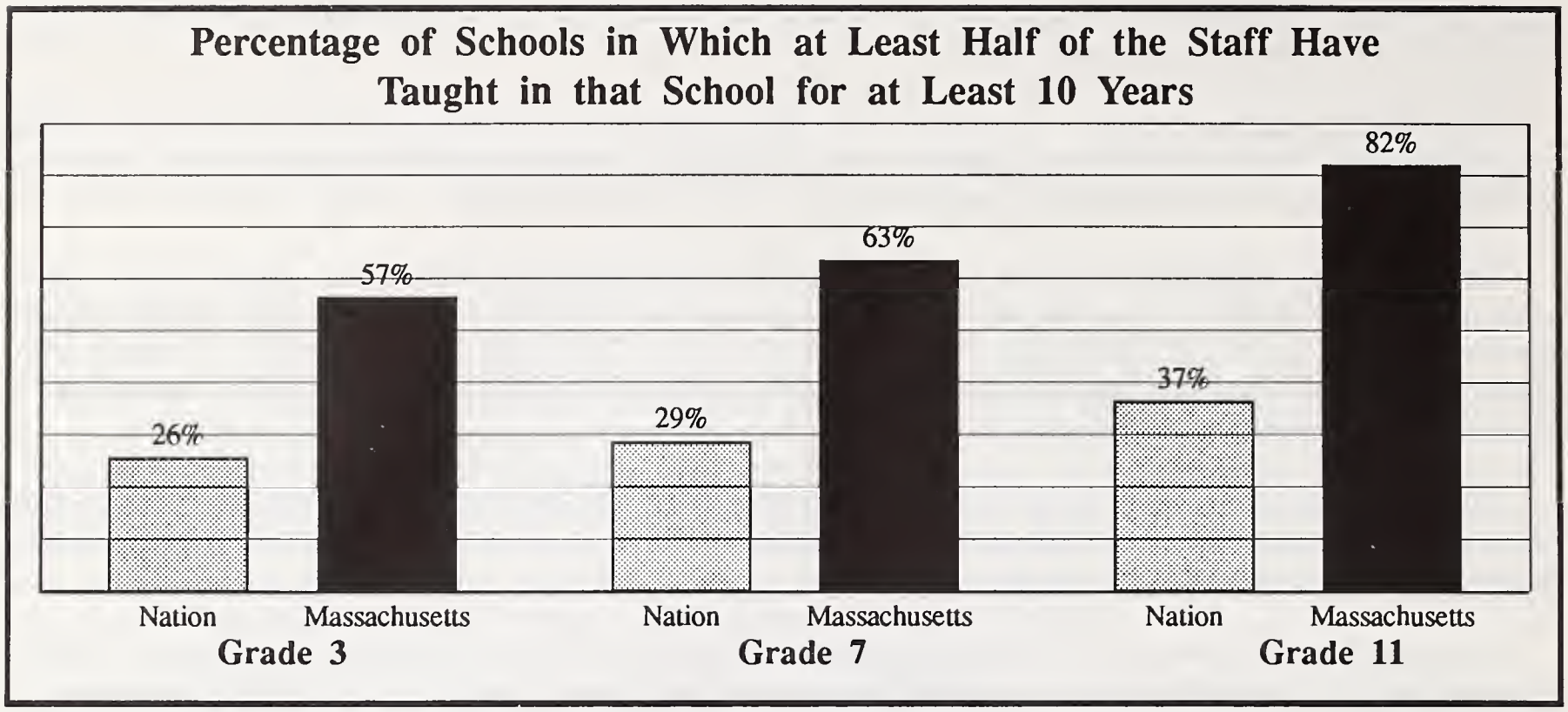
Several questions asked students about practices in their science classes. Practices in Massachusetts generally were very similar to those around the country. However, as was true for the seventh grade, Massachusetts students were much less likely to read from their science textbooks in class on a daily basis, but they also were less likely to ever do experiments or laboratory work by themselves.

Principal Questionnaires

As noted above, questionnaires were given to the principals in both assessments, and several questions in each were identical. The questions pertained to the principals themselves, their staffs, and practices within the school.

Perhaps one of the most important findings that comes from the national comparisons does not directly relate to student performance or even to curriculum. It is well known to most people involved with education in Massachusetts that the teaching force is growing older, and is not being replaced with younger teachers. The data collected from the principals brought home just how severe a problem this is, and how different Massachusetts is from the remainder of the country. In 57% of the schools in Massachusetts, a majority of the teachers have been teaching at that school for at least ten years; nationwide, the comparable figure is just 26 percent. At grade 7, the comparable results are 63 percent in Massachusetts versus 29 percent nationwide, and at grade 11, 82 percent versus 37 percent.

Typically, teaching staffs are analyzed in three groups: teachers with three or fewer years experience, teachers with between three and fifteen or so years experience, and those with more experience than that. The teachers in the most experienced group generally carry with them the traditions and historical perspective of the school, and pass that on to the second group. The second group generally are the innovators—the teachers who have been there long



enough to know what is going on, and just dissatisfied enough to try something new. The least experienced group usually looks to this second group for guidance and ideas on how to improve their teaching.

Nationwide, there is some cause for concern about where new teachers will look for leadership, because when new teachers join a faculty, they often find few teachers in this second group. This nationwide problem is exacerbated in Massachusetts by the effects of Proposition 2 1/2. When that proposition passed, many young teachers were laid off. To compound the problem, few new teachers have been hired since then.

Thus, today, Massachusetts finds itself with a situation that is a blessing for the short term, but has enormous potential negative consequences for the long term. The current teaching staff is probably the most experienced that Massachusetts ever has had. However, it is not clear what will happen in a few years when all these teachers approach retirement age at the same time. Will Massachusetts be able to replace them all at one time? If so, what kind of education will a uniformly inexperienced group of teachers, without the leadership that has traditionally come from the more experienced staff, be able to provide? An entire generation of teachers is largely missing in Massachusetts. It is clear that this is a problem that will have to be faced in the not-so-distant future.

The aging of staff is not true just for teachers. At all grades, Massachusetts principals were far more likely to have been head of their school for at least ten years. At grade 3, 47 percent of the Massachusetts principals have led their schools for at least ten years, as opposed to 25 percent of principals nationwide; at grade 7, the gap was 40 to 32; and at grade 11, 33 to 19.

Principals in Massachusetts schools, compared to schools nationwide, are also more likely to be males. At grades 3 and 11, the percentage of male principals in Massachusetts was fairly similar to national results (75 percent versus 72 percent at grade 3, and 92 versus 90 percent at grade 11), but the disparity at grade 7 was quite pronounced. Ninety-one percent of the seventh grade principals in Massachusetts were males, compared to 79 percent of such principals nationwide.

In spite of the higher level of experience of teachers in Massachusetts, responsibility for decisions regarding the selection of core instructional materials at grade 3 is distributed in Massachusetts very similarly to the way it is nationally. That is, in some districts in Massachusetts, such decisions are made by a group of teachers, and in other districts, such decisions are made at the district level. At the third grade level, both in Massachusetts and the nation, about an equal number of districts assign one of these two groups to be the primary decision maker. About 40 percent of the districts groups of teachers are the primary decision makers, and about another 40 percent of the decisions are made at the district level. The remaining 20 percent of the districts allow the primary decision maker to be the individual classroom teacher, a curriculum coordinator, the principal, or someone else.

While the same pattern holds true nationally at the seventh grade, that is not the case in Massachusetts. In Massachusetts, seventh grade teachers have much more influence on the selection of the core instructional materials. Either as individuals or in groups, in over 60 percent of the schools in Massachusetts teachers are the primary decision makers in this area.

The Massachusetts commitment to special education also was evident in the answers to the principal questionnaires. Massachusetts schools tend to have slightly more specialists working in their schools in all areas except special education, where they are far more likely to have several additional specialists. Nationwide, 30 percent of the schools at grade 3 have three or more special education teachers, while in Massachusetts 43 percent of the schools have that many. The advantage to Mssachusetts schools is even greater at the upper grades. In grade 7, 66 percent of the schools in Massachusetts have at least three special education teachers, compared to 40 percent nationwide; at grade 11, the figures are 73 percent for Massachusetts and 52 percent for the rest of the country.

APPENDIX A

A Brief Description of the Massachusetts Educational Assessment Program

Beginning in the spring of 1985, Department staff created four advisory committees of Massachusetts teachers to help determine the testing materials to be used with the new assessment program. One committee was formed for each content area (reading, mathematics and science) to provide direction for the development of the test content; a fourth committee's responsibility was to review all materials to be sure that they were fair and relevant to students of all ethnic, racial, and cultural backgrounds. Throughout that summer and the next fall, the committees met repeatedly to create tests that would be appropriate for use within the state. They used NAEP objectives and test questions as an initial guide, but expanded the scope of what was to be tested from that starting point. In the early fall, a survey was sent to all schools and districts in the state to apprise them of the committees' plans to date, and to receive feedback on those plans. Later that fall, the committees met to revise their plans based on the feedback, and to create the final versions of the tests.

A series of test administration workshops was conducted throughout the state between Febuary 24 and March 7, 1986. During the second week of March, testing materials were mailed to the schools, and between March 24 and April 11, 1986, the tests were administered.

The Massachusetts Educational Assessment Program tests were comprehensive, covering a broad range of objectives in reading, mathematics, and science.

The time required to take the full test battery and generate individual student test scores would have been between eight and twelve hours depending on the grade level. Because individual student scores were not required, however, it was not necessary to give each student the full test battery. Instead, the hundreds of test questions were divided among several forms of the test at each grade level, and each student completed only one form of the test. This procedure for testing, called matrix sampling, results in greater reliability and validity than does conventional testing, while keeping testing time to a minimum.

At each grade, students completed a questionnaire and one test form. Each test form contained three blocks. In order to provide national comparisons, the first block of each form was reproduced from a NAEP test, and administered under timed conditions paralleling those of NAEP. The remaining two blocks were untimed and contained additional NAEP items, as well as items developed for use on this test. Across the three grades, the entire battery consisted of 240 questionnaire items and 2,156 test questions, about two-thirds of which were taken from NAEP.

Chapter 188 required the testing of all students at the selected grade levels, with the exception of some special needs and limited English speaking students. Special needs students were to be excluded from the assessment program if so waived by their parents; limited English speaking students were to be excluded unless their parents requested otherwise. Across the three grade levels tested, a little over half the students in both of those categories were tested, and are included in this report. Of the students not excluded from testing for either of these reasons, over 97 percent were completely tested and are included in this report.

APPENDIX B

A Description of the Adaptation of NAEP to the Massachusetts Educational Assessment Program

The National Assessment of Educational Progress

The National Assessment of Educational Progress (NAEP) began in 1969 as a large-scale data collection effort to assess the state and progress of education in the United States. Since then, NAEP has assessed students many times in several different content areas. In the spring of 1986, NAEP assessed 9-, 13-, and 17-year olds, as well as students in grades 3, 7, and 11, in the areas of reading, mathematics, science and computers. The full battery of NAEP questions, including both questionnaire items as well as test questions, but excluding the subject area of computers, totalled 2,501 questions. These materials were administered to a total of 88,708 students in several hundred schools.

Differences between NAEP and the Massachusetts Assessment

No matter how much effort is put into making two assessments the same so that results will be comparable, differences of sufficient magnitude invariably exist that must be considered when interpreting the data. For example, Chapter 188 mandated that the results of the testing be made publicly available, school by school, which is a substantially different way of handling test results than is done with NAEP data. Thus, simple adherence to legislative requirements meant that comparisons between NAEP data and the Massachusetts data would not be direct or without a need for interpretation. The purpose of this appendix is to detail the differences between the two assessments, to explain what was done to minimize the effect of the differences, and to discuss the impact that the differences might have had on the results.

Reporting. As noted above, the reporting differences between NAEP and what was required for the Massachusetts assessment by Chapter 188 are substantial. NAEP does not release any test results below the level of region of the country, thereby guaranteeing anonymity to all participating schools, districts, and even states. On the other hand, it was known by all participants in the Massachusetts assessment, in advance of the testing, that results of that assessment would be publicly reported, school by school and district by district. The motivations of the two groups when tested, therefore, were considerably different. It can only be assumed that Massuchusetts students and administrators were motivated to do their best, and as a consequence, would score higher than the national sample, even if the achievement levels of Massachusetts students were no higher than their counterparts nationwide.

Population tested. Chapter 188 mandated which students would be tested and which students would be exempted from testing. These rules are different from the rules NAEP uses as to who will be tested, and thus the population of students eligible to be tested in Massachusetts was somewhat different from the national sample. NAEP's rules called for the testing of all students, except those "deemed unassessable by the school authorities," while in Massachusetts, Chapter 188 mandated the testing of all students, with limited exceptions.

One impact of the effect of local reporting now should be clear. With NAEP, a test administrator hired by NAEP was sent to the school with a list of students to be tested. That administrator reviewed the list with the principal of the school to be tested, and they jointly decided who should be tested and who should not. Note that in this case, the principal had just two motivations: to adhere to the guidelines for NAEP, and to serve the best interests of students who might be adversely affected by participating in the testing. In contrast, the Massachusetts

assessment had stricter guidelines on who would be tested (that is, all students would be tested unless specific criteria were met to exclude them), but local administrators had more motivation to exclude students who were not likely to score well on the tests. In Massachusetts, a very high percentage of students who were not special needs or bilingual were tested and therefore, it could be argued that Massachusetts performance was lowered because of that; however, it is not known how many special needs and bilingual students who could have been tested, and therefore would have been tested for NAEP, were excluded from the Massachusetts testing.

A second difference between the groups tested in the two assessments that should have had no impact on results, but did allow for some margin of difference, is the fact that the Massachusetts assessment was census testing (that is, every student in every school was to be tested), while NAEP uses a sampling design (that is, a sample of students in a sample of schools is tested). NAEP uses a complex set of weights to estimate national performance based upon the sample they have drawn, but there always remains some element of doubt about the actual implementation in the field of sampling design.

Test format. The NAEP budget for two years of testing was $8 million; the Massachusetts assessment, which involved the testing of more students in more schools, was under $1 million. Obviously, some changes had to be made in the Massachusetts assessment to keep costs down. Some of these changes probably had some impact on the direct comparability of the results.

A major difference was the test materials themselves. Because of the volume of the Massachusetts assessment, it was important to keep down the number of pages included in each booklet. Thus, while NAEP typically puts three or fewer test questions on one page of text, the Massachusetts assessment test booklets usually were double columned, and contained three to six questions per page. Also, while the Massachusetts assessment tried to use pictures and charts exactly as used by NAEP, sometimes the size of them had to be changed, and sometimes the clarity of reproduction was not the same. NAEP had students respond to questions directly in the answer booklet, while at grades 7 and 11, Massachusetts had students record their answers on separate answer sheets. In contrast to the concerns raised in the previous section, there is considerable literature about the effects these changes would have made. In every case, they put Massachusetts students at a disadvantage, and thus, tended to lower the performance of those students relative to their counterparts nationwide.

Use of trained administrators. NAEP sent staff to every school participating in the testing to oversee the test administration. Massachusetts mailed all their materials to schools, and test administration was a local concern. To be sure, training workshops were conducted at the six regional education centers in the state in order to properly prepare staff for the administration of the tests, and manuals detailing the administration procedures were included with all the testing materials, but there was no requirement that the workshops be attended or the manuals read. The impact of this difference on test results is unknown.

Timed tests. Some testing procedures that were appropriate for NAEP, given the intent of that program, had to be modified for the Massachusetts assessment. For example, the NAEP tests were timed. Each student was given a short period of time to answer a series of attitudinal questions as well as test questions. It was not unusual for there to be forms which were completed by fewer than half the students taking them. Given the direct relationship between the testing and curriculum evaluation that was a desired outcome of the Massachusetts assessment, that was considered unacceptable, and so students were given all the time they needed to complete the test for all parts except the first block. Even in that first block, which was a virtual duplication of NAEP materials, the attitudinal questions were left off in order to provide students with a better chance of completing the test, and thereby muting the concerns that surely would have arisen from school staff if more students had not completed.

Use of "I don't know." Another difference between the two programs revolves around guessing and related issues. Consistent with its mission, NAEP frequently uses the distractor, "I don't know," with its multiple choice questions. While that answer option makes great sense for NAEP, it provides a confusing message to students taking the Massachusetts assessment. Since results are reported locally in Massachusetts, and since answering "I don't know" clearly is an incorrect answer to a test question, students in Massachusetts should never select that option when it is presented. While some did, it can only be assumed that administrators must have discouraged them from doing so. Also, NAEP provides no instructions to students as to what to do if they don't know the answer to a question.

Again, given the local reporting requirements of the Massachusetts assessment, it was necessary to provide some direction about guessing so that tests would be administered uniformly statewide. The Massachuetts assessment directions therefore told students to "make your best guess," while the NAEP directions said nothing about guessing.

Attempts to Make the Data As Comparable As Possible

Given all the issues raised above about the comparability of the data from the two assessments, it was necessary to plan carefully to make the data as comparable as possible. The purpose of this section is to describe those plans and the changes that were made in them as the assessment progressed.

As noted above, the basic source of information about the performance of students in Massachusetts relative to students nationwide was to come from the first block of each form. In each case, this block was a duplication, to the extent possible, of a NAEP block. Certain changes had to be made in order to meet the administrative requirements of the Massachusetts tests. For example, NAEP had several questions that required a short written response ("open-ended questions") rather than being multiple choice. The administrative restrictions on the Massachusetts tests did not permit switching item types within a block. Rather than just discarding the open-ended questions, Massachusetts kept the blocks as comparable as possible by converting the open-ended questions to multiple choice. No NAEP-comparable data could be obtained, of course, once the item had been changed that much, but by including the item, the context in which students were taking the multiple choice questions around those open-ended questions would remain as similar as possible. In the case of one open-ended question at grade 3, it was felt that no multiple choice item could replace it, so it was left as an open-ended question. Students wrote their answer to the question right in the test booklet, but it was not scored. The only role that question played was making the context for the questions around it as correct as possible.

Two major substudies also were conducted by including test questions in the assessment materials that would improve the comparability of future assessments. Most NAEP blocks contained the option "I don't know," which, as noted above, was certainly interpreted differently by students in Massachusetts (where local reporting was part of the assessment) than by students in the NAEP sample. In order to attempt to eliminate the "I don't know" option in future assessments, several NAEP blocks were duplicated with that option deleted. If a predictable effect of eliminating that option could have been found, future editions of the Massachusetts assessment could be published without the option, and appropriate adjustments to the data made. Unfortunately, attempts to find that predictable effect have been unsuccessful to date.

The other major substudy was far more successful. As noted above, the test items used to make the comparisons to NAEP were those in intact blocks—sections of test materials that were directly taken from NAEP and had no questions specific to Massachusetts. In addition, several other NAEP items were included in the test, scattered thoughout the forms. There turned out to be no significant differences between comparisons made on the "intact" items versus the "scattered" items. This finding will open up substantially future editions of the Massachusetts assessment tests.

Another difference between NAEP and the Massachusetts assessment was the use by Massachusetts of separate answer sheets on which students recorded their answers. In order to have students make as few "bookkeeping" errors as possible, the answer sheets were custom designed for each form so that the correct number of questions and the correct number of options for each question were printed on the answer sheet for each student. Since each answer sheet matched up to only one booklet, the procedure also helped to assure that the number of the form the student was taking was correctly recorded on the answer sheet.

The final modifications in the plans to maximize the comparability of the data came after receiving the NAEP data tapes. Through those, it was confirmed that many students did not finish the NAEP blocks, and thus, students in Massachusetts had a major advantage in taking the test. In order to minimize this problem, only test questions which had been reached by at least 95 percent of the NAEP sample were used in making comparisons. This resulted in discarding a great deal of information that was available, but it was felt that any attempts to interpret results on questions that were not reached by large numbers of students would be too fraught with error.

Another refinement of the NAEP data was the use of Block 1 data only. When intact NAEP blocks were used by Massachusetts, they were used as the first section of questions only, whereas NAEP typically administered each block an equal number of times in each of the three sections of the test. To obtain comparable data, NAEP data were used only when students were public school students, enrolled in the correct grade, and took the block under consideration as the first section in their booklet.

Missing Reading Results

At the present time, NAEP has not released its data for reading. Consequently, comparisons to national performance can be made in mathematics and science only.

A Final Caution

All the differences between NAEP and the Massachusetts assessment are potential causes of systematic error. As shown above, a variety of methods was used to minimize systematic error. In addition to the systematic error, however, is added random error, caused by the fact that both assessments used sampling of students, rather than administering all test questions to the entire population.

The size of the random error is of particular concern with the NAEP data. While NAEP administered each test question to approximately 2,500 students nationwide, this is a total number, summed across the three sections of the test. As noted above, students scored significantly better on test questions when they were taken in an early section rather than a late one, and in order to match NAEP data most closely, Massachusetts data are compared only to the NAEP data in the same section. Thus, two-thirds of the NAEP data are of no value for comparisons, and the NAEP sample is reduced to approximately 800 students per item.

Because of matrix sampling, the Massachusetts assessment data also have some random error associated with them, but the error is much smaller than that associated with the NAEP data. Each item in Massachusetts was taken by approxiamtely 4,000 students.

Taken together, differences on performance of individual items of less than four percentage points should be considered not statistically significant. On collections of items (for example, the average of all items at a particular grade level), differences of less than two percentage points should be considered not statistically significant. The statistic of interest in this report is the difference between national and Massachusetts performance. Because the random errors of both assessments occur independently, the error associated with the differences is equal to the sum of the two. In short, any difference between NAEP and Massachusetts results should be read as an accurate statement within four percentage points. That is, suppose the difference between national and Massachusetts performance on a given item was six percentage points. It would be more correct, given the potential for random error associated with those observations, to refer to the difference as being in the range of two to ten percentage points, rather than interpreting it as an unequivocal difference of six percentage points.

Note that this is true for the interpretation of individual items only. The averages of collections of items from the two assessments, given as they are to several samples of students, have much smaller random errors associated with them. Those error margins range from two percentage points at grade 3, where a more limited number of items was available for national comparisons, to under one percentage point at grade 11.

MASSACHUSETTS DEPARTMENT OF EDUCATION

REGIONAL CENTERS

Micki Siegel
Northwest Regional Center
Mark Hopkins Hall
Church Street
North Adams, MA 01247
413/664-4511

Paul Burnim
Greater Springfield Regional Center
Macek Drive
Chicopee, MA 01013
413/594-8511

Dorothy Earle
Northeast Regional Center
219 North Street
North Reading, MA 01864
617/664-5723

Laurie Slobody
Central Mass. Regional Center
Beaman Street, Route 140
West Boylston, MA 01853
617/835-6266

Peter Cirioni
Greater Boston Regional Center
75 Acton Street
Arlington, MA 02174
617/641-4870

Pat O'Brien
Southeast Regional Center
P.O. Box 29
Middleboro, MA 02346
617/947-1231

Related publications about the Massachusetts Educational Assessment Program:

- *The Massachusetts Educational Assessment Program: A Summary of the First Year's Statewide Results*
- *Using the School Reports*
- *Reading and Thinking: A New Framework for Comprehension*
- *Science in the Elementary Schools*
- *The 1987 Massachusetts Educational Assessment Program: Description of Test Content and Reporting Categories*
- *Moving Geometry From the Back of the Book: A Report on Geometry and Measurement in the 1986 Assessment*

Advanced Systems IN MEASUREMENT & EVALUATION, INC.
MOSES PAUL HOUSE • 114 LOCUST STREET, DOVER, NEW HAMPSHIRE 03820